Developing a question answering (QA) system for spoken lectures.

**Challenges:**
1. ASR transcriptions contain recognition errors.
2. Lecture videos can be long.
3. Answers can be longer than couple of words.
4. Task is domain specific.

**Proposed Approach:**
Convert the problem into machine comprehension style QA.

- **Machine Comprehension:** Textual questions + Short passages
- **Use competitive end-to-end neural network models.**
- **Split lecture transcriptions into short pseudo passages.**
- **Automatically match questions with passages.**

**Spoken Lecture Processing System**

- **Lecture Videos:** Short videos (~5 minutes long)
- **ASR System:** Acoustic model (AM) trained with Kaldi\([1]\) using 2.7 hours lectures.
- **Lecture Transcriptions:** 4 different engineering courses (totally ~4 hours).
- **QA System:** End-to-end neural network model.

**QA Example**

**Paragraph:**

Now just as with the Fourier transform there are a number of properties of the Laplace transform that are extremely useful in describing and analyzing signals and systems. For example one of the properties that we in fact took advantage of in our discussion last time was the linearity property which says in essence that the Laplace transform of the linear combination of two functions is the same linear combination of the associated Laplace transforms. Also there is a very important and useful property which tells us how the derivative of a time function rather the Laplace transform of the derivative is related to the Laplace transform in particular the Laplace transform of the derivative is the Laplace transform x(t) multiplied by s and as you can see by just setting s equal to omega in fact this reduces to the corresponding Fourier transform property.

**Question 1:** What is the linearity property in the Laplace Transform?

**Question 2:** How is the Laplace transform of the derivative of a time function related with the Laplace transform of this time function?

**Method**

**MatchLSTM with Answer Pointer\([6]\):**

1. **Preprocessing:** Representations for passage P and question Q.
   \[ H^p = LSTM(P), H^q = LSTM(Q) \]

2. **MatchLSTM:** Question aware passage representations \( H^q \).
   \[
   \begin{align*}
   \text{\overline{c}_i} &= \tanh(W^w \text{\overline{h}}^p_i + (W^b h^b_i + W^e h^e_i) \otimes e_Q) \\
   \text{\overline{r}_i} &= \text{softmax}(W^r \text{\overline{h}}^p_i + b \otimes e_Q) \\
   \text{\overline{f}_i} &= [h^p_i, \overline{c}_i] (\text{attention vector})
   \end{align*}
   \]
   \[
   \overline{h}_i^q = \text{LSTM}(\overline{f}_i, \overline{h}_i^{q-1})
   \]
   \[
   H^q = [\overline{h}_1^q, ..., \overline{h}_T^q] (\text{same calculations in reverse order.})
   \]
   \[
   H^q: \text{concatenation of } H^q \text{ and } \overline{h}_T^q
   \]

3. **Answer Pointer:** Prediction of the answer (span of words).

**Passage-Question Relevancy:** Assign questions to pseudo passages.

- **Split transcription into fixed length pseudo passages.**
- **Compute relevance scores between pseudo passages and questions using question aware passage representations.**
- **Select a single passage (with the maximum score) for each question.**

**H_{rel}**: concatenation of \( h_i^p \) and \( \overline{h}_i^q \)

\[
D = \tanh(W^d h_i^p h_i^q + b_i^d) \quad \theta = \text{sigmoid}(W^\theta D + b_{\theta i}^d)
\]

**Conclusion and Future Work**

- **A QA system (based on competitive machine comprehension models) was developed for spoken lectures in the signal processing area.**
- **A passage-question matching stage was proposed to handle a realistic scenario where the answer for each question is searched in a chapter of the course lectures.**
- **Future Work:** Extend the data to increase diversity. Deal with unanswerable questions in QA system.

**Experiments**

**Train-Test Scenarios:**

- **short-short:** Short (manually split) passages for train and test data.
- **short-long:** Short passages for train data + long passages (transcription of the whole video) for test data.
- **long-long:** Long passages (~650 words) for train and test data.
- **window-window:** Fixed length (200 words) pseudo passages for train and test.

**WERs:**

- GSM-s(i) (13.0%), GSM-sa (10.5%), DNN (6.7%).

**QA systems**

- **Train-Test Scenarios:**
  - **Test Set F1 Score**
    - **with known passage-question pairs**
      - GSM-s(i)
      - GSM-sa
      - DNN
    - **Ref.**
      - short-short 56.38
      - short-long 23.39
      - long-long 27.84
      - window-window 38.76

**References and Acknowledgements**


This work is supported by the TUBITAK ARDEB 3501 Program (Project No: 117E202).